

長岡技術科学大学大学院  
工学研究科修士論文

題 目

独立部分空間分析による  
音源分離の精度向上に関する研究

指導教員 準教授 杉田 泰則

著者 電気電子情報工学専攻  
14318486 増田 翔太

提出期日 平成30年2月9日

# 目 次

<b>第 1 章 序論</b>	<b>1</b>
1.1 背景と目的 . . . . .	1
1.2 本論文の構成 . . . . .	2
<b>第 2 章 頭部伝達関数と音源方向の推定</b>	<b>3</b>
2.1 座標系 . . . . .	3
2.2 頭部伝達関数 . . . . .	4
2.3 頭部伝達関数による音源方向の推定 . . . . .	5
2.4 まとめ . . . . .	7
<b>第 3 章 提案手法の基礎となる手法</b>	<b>9</b>
3.1 独立部分空間分析 . . . . .	9
3.2 方向推定によるスペクトルの分類 . . . . .	11
3.3 まとめ . . . . .	12
<b>第 4 章 提案手法</b>	<b>13</b>
4.1 複素誤差の絶対値 . . . . .	13
4.2 クロススペクトル . . . . .	14
4.3 まとめ . . . . .	15
<b>第 5 章 シミュレーション</b>	<b>16</b>
5.1 実験条件と評価指標 . . . . .	16
5.1.1 実験条件 . . . . .	16
5.1.2 評価指標 . . . . .	18
5.2 実験結果 . . . . .	18
5.2.1 case1 . . . . .	19

5.2.2 case2 . . . . .	20
5.3 考察 . . . . .	23
5.4 まとめ . . . . .	25
<b>第 6 章 結論</b>	<b>26</b>
<b>付録 A シミュレーション結果</b>	<b>27</b>
A.1 シミュレーション case1 の結果 . . . . .	27
A.2 シミュレーション case2 の結果 . . . . .	32
<b>謝辞</b>	<b>36</b>
<b>参考文献</b>	<b>38</b>

# 第1章 序論

本章では、本論文の研究背景および目的、本論文の構成について述べる。

## 1.1 背景と目的

混合音源から特定の音を抽出する技術を音源分離と呼び、補聴器、音声認識、遠隔会議、音声強調やロボットなどに応用できることから広く研究されている。音源分離技術として最も基本的な技術は、マイクロホンを複数用いるマイクロホンアレー技術である。しかしこの技術は音源の到来方向が事前情報として必要であり、到来方向が未知である場合には十分な分離性能を得ることができない。そこで到来方向を必要としない主な手法として、周波数マスキングや独立成分分析が挙げられる[1]。周波数マスキングは同時刻、同周波数帯において、2つ以上の音源のスペクトルが重ならないというスパース性を仮定し、観測音源のスペクトルから目的音源以外のスペクトルを取り除くことで音源分離を実現する手法である[2, 3]。また、独立成分分析(ICA, Independent Component Analysis)は音源の統計的な独立性を仮定し、音源の独立性ができる限り大きくなるように元の音源を推定することで実現できる[4, 5]。しかし独立成分分析による音源分離は、原理そのままだと観測音源の数と同じ数の音源までしか分離することができない[6]。そこでICAを拡張した手法として、独立部分空間分析(ISA, Independent Subspace Analysis)が存在する[6]-[5]。ISAでは複数の信号の集合である部分空間同士の独立性は仮定するが、部分空間内に存在する信号同士の独立性は仮定しないため、1つの混合音源のスペクトログラムからでも音源分離を実行できる[8]。

補聴器への応用を考えると、少ない数のマイクで観測した音源からマイク数以上の音源を分離できる手法や、分離音源の到来方向が推定できる手法が重要である。

ところで、ヒトは2つの耳のみで音源分離を実現している。これは音源が頭部伝達関数(HRTF)と呼ばれるヒトの耳に到来した際に頭部や胴体に起因する伝達関数の影響を受け、左右の耳に到達した音源間で位相差(IPD)や強度差(ILD)が生じるためである[9, 10]。ヒトはこれらの情報を用いて音源分離を実現している。

ヒトの音源分離の原理と同様に、HRTFを利用した音源分離手法を両耳聴モデルと呼ぶ。西川は独立部分空間分析(ISA)に両耳聴モデルを適用した手法を提案し、正中面を含む到来方向の推定と音源の分離を同時に実現した[14]。これは、ISAを用いて分離した音源のス

ペクトルから到來方向を推定するための特徴量を取り出し, データベースの角度ごとの特徴量と比較して最も近い角度にスペクトルを分類することにより実現される. 従来の手法では角度ごとの特徴量として観測音源の左右のパワーの強さにより重み付けされた IPD を用いていたが, 従来より角度ごとの特徴が出やすい重みと特徴量を用いることで分離精度が向上すると考えられる.

そこで本論文では, 混合音源から推定したスペクトルに含まれる角度ごとの特徴量として複素誤差の絶対値を, データベースとの比較の際に用いる重みとしてクロススペクトルを提案する. 複素誤差の絶対値を採用することにより IPD では切り捨てられていた音圧レベルの情報を利用することができ, データベースとの比較精度が向上すると考えられる. また, 重みにクロススペクトルを採用することにより, 観測音源の左右のパワーの平均では排除できなかった, 左右のどちらかの音源のパワーのみが大きい周波数を排除し, 左右の音源で共通してパワーが大きい周波数のみを選択することができる. これにより雑音の少ない周波数スペクトルでのデータベースとの比較が容易になり, 分離精度が向上すると考えられる. シミュレーションでは, 従来手法と提案手法の角度ごとの特徴量と重みの組み合わせを変えてそれぞれ分離精度の比較を行い, 提案手法同士の組み合わせが最も分離精度に優れていることを示す.

## 1.2 本論文の構成

本論文の構成は以下のとおりである. 第 2 章では, 頭部伝達関数とそれを用いた音源方向の推定について説明する. 第 3 章では, 提案手法の基礎となる手法である独立部分空間分析と, その手法における方向推定によるスペクトルの分類の処理について説明する. 第 4 章では, スペクトルの分類において用いられている角度ごとの特徴量と重みについて, 従来の IPD と左右のパワーのスペクトログラムの平均に代わるものを探求する. 第 5 章では, シミュレーションにより, 提案手法と従来手法の分離性能の比較を行い, 提案手法が優れていることを確認する. 最後に第 6 章で本論文の結論を述べる.

## 第 2 章 頭部伝達関数と音源方向の推定

本章では、まず本論文で用いる座標系を定義し、その後ヒトの頭部伝達関数による音源方向の推定の仕組みについて説明する。

### 2.1 座標系

音源の方向を表すために、本論文では図 2.1 に示す球座標系を用いる [10]。この座標系での原点は、ヒトの両外耳道入り口を結ぶ線分の中点である。また、右眼窩点と左右の耳珠を含む平面を水平面と呼び、両外耳道入り口を通り水平面に垂直な面を横断面と呼ぶ。最後に、水平面と横断面のどちらにも直交する平面を正中面と呼ぶ。また、角度  $\theta$  は方位角、角度  $\phi$  は仰角と呼ばれている。

本論文では基礎的な検討を行うために仰角  $\phi$  は  $0^\circ$  のみとし、これ以降  $\phi$  は省略する。よって、正中面は  $\theta = 0^\circ$  と  $\theta = 180^\circ$  の 2 点となる。

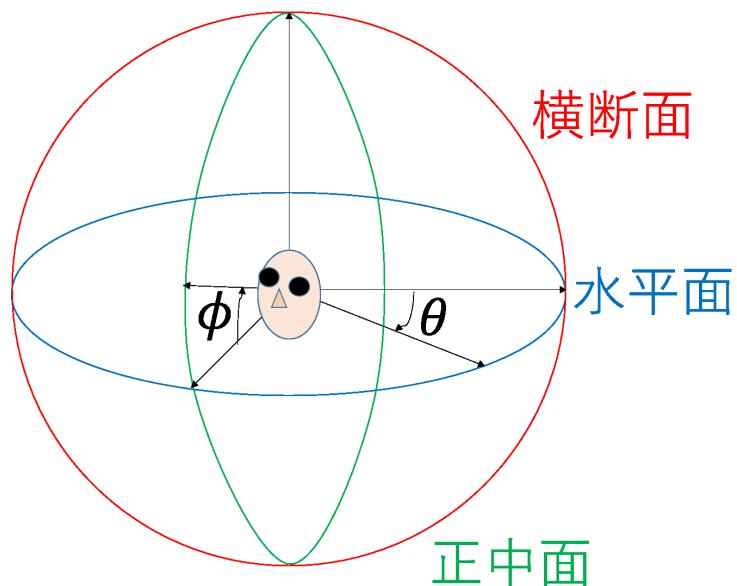


図 2.1 球座標系

## 2.2 頭部伝達関数

音がヒトの耳に到来するとき、胴体や頭、耳の形などによる反射や回折が生じる。頭部伝達関数 (HRTF, Head-Related Transfer Function) はこれらによる伝達特性を表したものであり、次式で定義される [10].

$$H_L(\omega, \theta) = \frac{G_L(\omega, \theta)}{F(\omega, \theta)} \quad (2.1)$$

$$H_R(\omega, \theta) = \frac{G_R(\omega, \theta)}{F(\omega, \theta)} \quad (2.2)$$

ここで、 $H_L(\omega, \theta)$  と  $H_R(\omega, \theta)$  は左右の HRTF で、 $\omega$  は周波数である。 $G_L(\omega, \theta)$ ,  $G_R(\omega, \theta)$  は音源から各耳の外耳道入り口までの伝達関数である。また、 $F(\omega, \theta)$  は各音源から頭部中心位置までの頭部がない場合の伝達関数である。次のように HRTF を逆フーリエ変換すると頭部インパルス応答 (HRIR, Head-Related Impulse Response) が得られる。

$$h_L(t, \theta) = FT^{-1}\{H_L(\omega, \theta)\} \quad (2.3)$$

$$h_R(t, \theta) = FT^{-1}\{H_R(\omega, \theta)\} \quad (2.4)$$

ここで、 $t$  は時間、 $FT^{-1}\{\cdot\}$  は逆フーリエ変換を表す。

音源  $s(t)$  がヒトの耳に到来するとき、HRIR が畳み込まれる。HRIR が畳み込まれた音源はバイノーラル音源と呼ばれ、畳み込み演算\*を用いて次式のように表す。

$$s_L(t) = h_L(t, \theta) * s(t) \quad (2.5)$$

$$s_R(t) = h_R(t, \theta) * s(t) \quad (2.6)$$

ヒトはバイノーラル音源から得られる情報を手がかりにして音源方向の推定を行っている。このとき、左右方向と前後方向で異なる情報を用いて方向推定を行っている。ここで図 2.2 に示すように、左右方向とは正中面からの角度、前後方向とは横断面からの角度のことである。

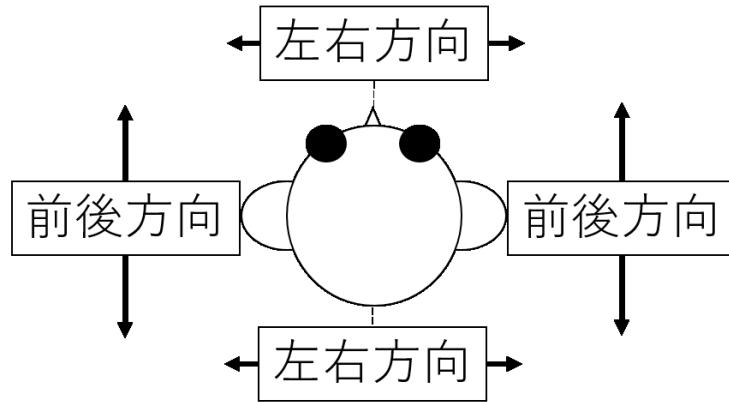


図 2.2 左右方向と前後方向

### 2.3 頭部伝達関数による音源方向の推定

左右の耳で観測された音源  $x_L(t)$ ,  $x_R(t)$  には到来方向に応じた位相差と強度差が生じており, ヒトはこれらを左右方向の推定の手がかりとしている. この位相差は両耳間位相差 (IPD, Interaural Phase Difference), 強度差は両耳間強度差 (ILD, Interaural Level Difference) と呼ばれており, IPD  $\psi(\omega, \theta)$  と ILD  $\xi(\omega, \theta)$  はそれぞれ次の式で定義される.

$$\psi(\omega, \theta) = \arg \left\{ \frac{H_R(\omega, \theta)}{H_L(\omega, \theta)} \right\} \quad (2.7)$$

$$\xi(\omega, \theta) = 20 \log \left| \frac{H_R(\omega, \theta)}{H_L(\omega, \theta)} \right| \quad (2.8)$$

ここで,  $\arg\{\cdot\}$  は位相角を表す. 各周波数における IPD と ILD を図 2.3 に示す. ただし, 本論文ではマサチューセッツ工科大学で配布されている HRTF データベース [11] を利用する.

ヒトは音源方向の推定のために主に 1.5kHz 未満では IPD, それ以上では ILD を用いている. これは以下の理由であると考えられている.

1. 位相差は周波数が高いほど大きくなるが,  $-\pi$  から  $\pi$  の範囲外となると多義性が生じ, 方向を推定できなくなる. 図 2.3(a) において 800Hz の IPD では方位角  $\theta = \pm 60^\circ$  付近でこの範囲を超えたことがわかる.
2. 音の周波数が高くなるほど回折が起きづらくなり ILD は大きくなる. 図 2.3(b) からもこの傾向が読み取れる.

一方で、IPD と ILD により前後方向を推定することは困難である。 $\theta=0^\circ$  から  $\theta=180^\circ$  の IPD と ILD を図 2.4 に示す。同図から、前後方向において IPD と ILD が同一の値を取る角度が存在することがわかる。IPD と ILD に多義性が生じ、音源の到来方向を一意に決定することができない。そこで、ヒトは HRTF の振幅特性に起因するピークとノッチにより前後推定を行っている。

$\theta=0^\circ$  と  $\theta=180^\circ$  における HRTF の振幅特性を図 2.5 に示す。HRTF の振幅特性にはピークとノッチが存在しており、ヒトはこれらを検出することにより前後方向の推定を行っている。具体的には、4kHz 付近のピーク (P1)、および 4kHz 以上における 1 つ目と 2 つ目のノッチ (N1, N2) を利用しており、これらはスペクトラルキュ [12] と呼ばれている。図 2.5 から分かるように  $\theta=0^\circ$  と  $180^\circ$  において、N1 と N2 の周波数位置が異なっているため、音源到来方向の推定に用いることが可能となる。一方で P1 の位置はどの角度においてもほとんど同じ周波数位置に存在する。これにより P1 はヒトの聴覚システムにおいてノッチ N1, N2 を探索するリファレンス情報として用いられていると考えられる。

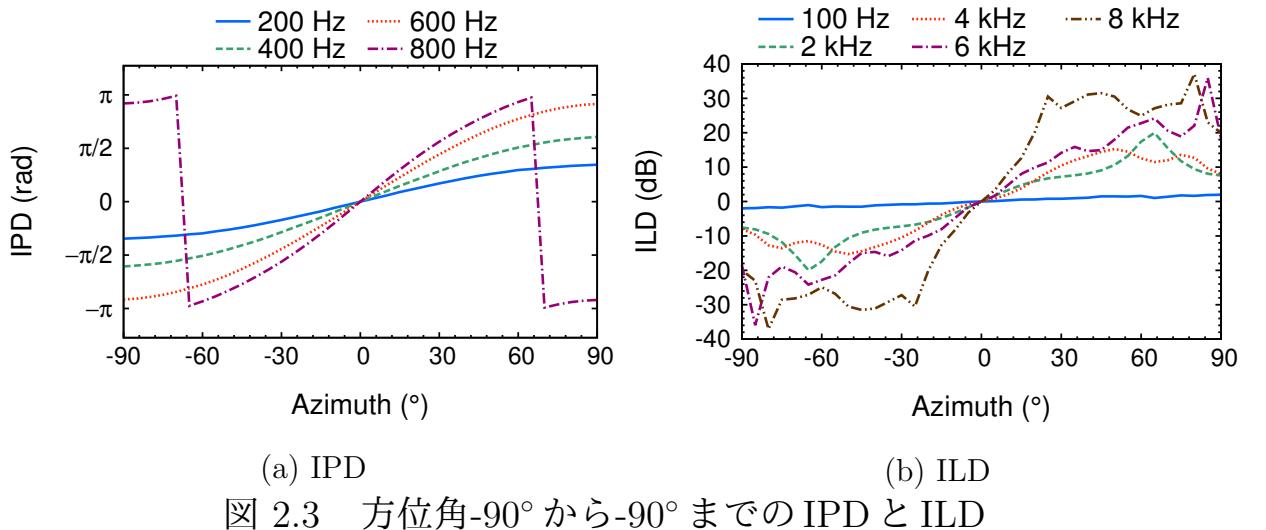


図 2.3 方位角-90°から-90°までのIPDとILD

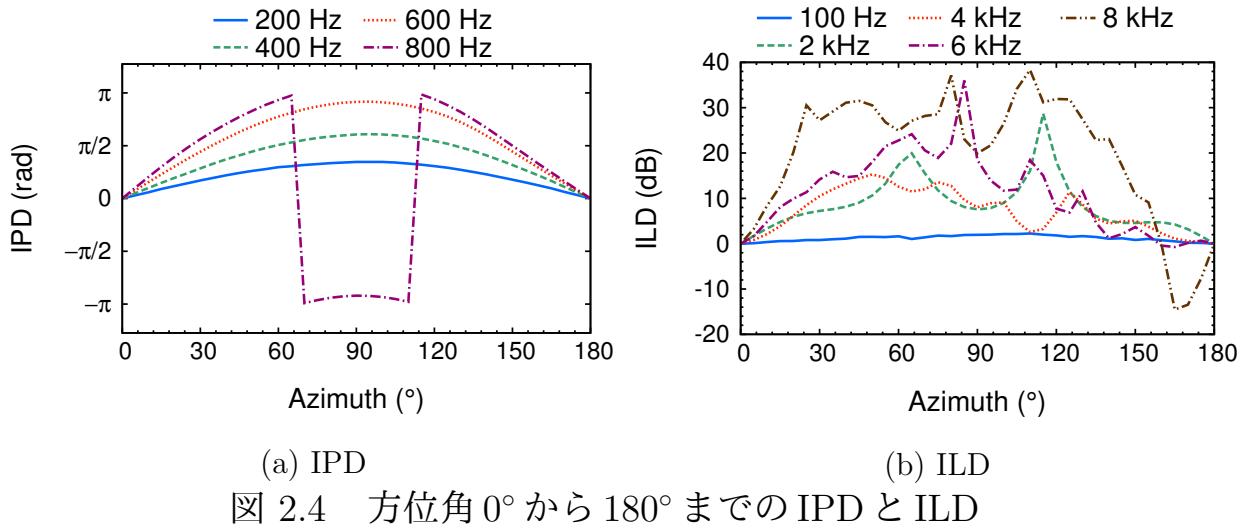


図 2.4 方位角  $0^\circ$  から  $180^\circ$  までの IPD と ILD

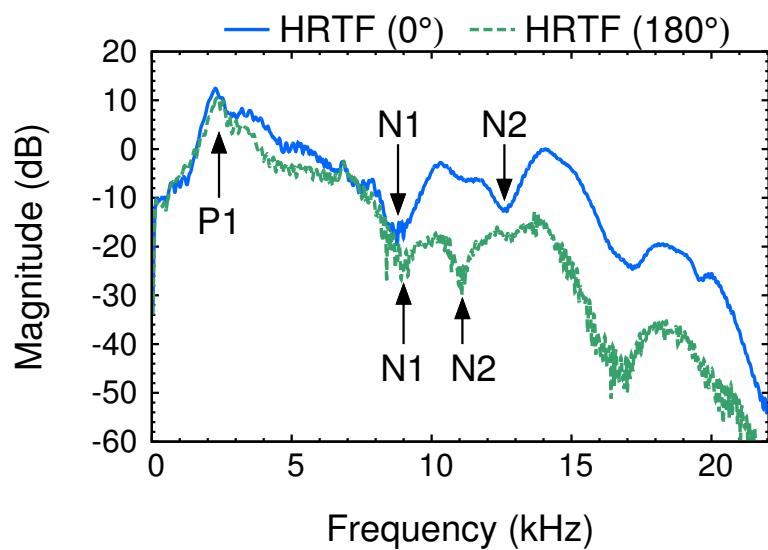


図 2.5 方位角  $0^\circ$  と  $180^\circ$  におけるスペクトラルキュー

## 2.4 まとめ

本章ではまず本論文で用いる座標系について定義した。次にヒトの頭部伝達関数による聴覚機能について説明し、ヒトは左右方向の推定に、1.5kHz未満の周波数では IPD を、それ以上の周波数では ILD を用いていることを説明した。また、前後方向の推定では IPD や ILD が同一の値を取る角度が存在するため、スペクトラルキーを用いていることを説明した。

# 第3章 提案手法の基礎となる手法

本章では提案手法の基礎となる独立部分空間分析を用いた分離技術について述べる。独立部分空間分析と独立部分空間分析を用いた音源分離について述べた後、独立部分空間分析を用いた音源分離における方向推定とスペクトルの分類の処理について詳しく述べる。

## 3.1 独立部分空間分析

独立部分空間分析 (ISA, Independent Subspace Analysis) とは独立成分分析 (ICA, Independent Component Analysis) を拡張した手法である [6]。ICA はすべての信号が独立であると仮定するが、一方で ISA では複数の信号の集合 (部分空間) の独立性を仮定し、同じ部分空間に属する信号の独立性を仮定しない。そのため ICA と同じアルゴリズムを用いて信号を推定することができるが、後処理として推定した信号を部分空間ごとに分類する必要がある。

Casey and Westner は観測音源のスペクトログラムに対して ISA を適用することにより、音源のスペクトルを推定し、1つの観測音源から音源分離を行う手法を提案した [13]。西川はこれを両耳聴モデルに応用し、さらに推定される成分の順序が不定である問題を解決した手法を提案した [14]。以下にその手順を示す。

$P$  個の音源  $s_i(n)$  ( $i = 1, 2, \dots, P$ ) がヒトの頭部に向かって方位角  $\theta_i$  から到来しとき、左右の観測音源  $x_L(n), x_R(n)$  は、次のようにバイノーラル音源の重ね合わせで表現される。

$$x_L(n) = \sum_{i=1}^P s_{L,i}(n) \quad (3.1)$$

$$x_R(n) = \sum_{i=1}^P s_{R,i}(n) \quad (3.2)$$

ここで、 $s_{L,i}$  と  $s_{R,i}$  は  $i$  個目のバイノーラル音源で、次のように音源  $s_i$  に頭部インパルス応答 HRIR が畠み込まれたものである。

$$s_{L,i}(n) = h_L(n, \theta_i) * s_i(n) \quad (3.3)$$

$$s_{R,i}(n) = h_R(n, \theta_i) * s_i(n) \quad (3.4)$$

左右の観測音源  $x_L(n)$ ,  $x_R(n)$  のスペクトログラムは次のように短時間フーリエ変換(STFT, Short-Time Fourier Transform)により得られる.

$$X_L(k, m) = STFT\{x_L(n)\} \quad (3.5)$$

$$X_R(k, m) = STFT\{x_R(n)\} \quad (3.6)$$

ここで,  $STFT\{\cdot\}$  は STFT を表し,  $k(=0, 1, \dots, K-1)$  は周波数インデックス,  $m(=0, 1, \dots, M-1)$  は時間インデックスである.

ISA では  $i$  個目の音源のスペクトログラム  $X^{(i)}(k, m)$  が  $\rho_i$  個のスペクトル  $Z_j^{(i)}(k)(j=1, 2, \dots, \rho_i)$  の荷重和で表すことができると考える. よって以下の式のようになる.

$$X^{(i)}(k, m) = Y_1^{(i)}(m)Z_1^{(i)}(k) + Y_2^{(i)}(m)Z_2^{(i)}(k) + \dots + Y_{\rho_i}^{(i)}(m)Z_{\rho_i}^{(i)}(k) \quad (3.7)$$

$$= \sum_j^{\rho_i} Y_j^{(i)}(m)Z_j^{(i)}(k) \quad (3.8)$$

$Y_j^{(i)}(m)$  は時変係数である.  $i$  個目の音源のスペクトログラム  $X^{(i)}(k, m)$  は  $Z_j^{(i)}(k)$  の集合(部分空間)であり, この部分空間が独立となるよう推定を行うことで音源のスペクトログラムを推定できる.

ISA を左右の耳で観測される  $i$  個目の音源のスペクトログラム  $X_L^{(i)}(k, m)$ ,  $X_R^{(i)}(k, m)$  に対して同時に用いることで次式を得る.

$$X_L^{(i)}(k, m) = \sum_{j=1}^{\rho} Y_j^{(i)}(m)Z_{L,j}^{(i)}(k) \quad (3.9)$$

$$X_R^{(i)}(k, m) = \sum_{j=1}^{\rho} Y_j^{(i)}(m)Z_{R,j}^{(i)}(k) \quad (3.10)$$

ここで  $Z_{L,j}^{(i)}(k)$  と  $Z_{R,j}^{(i)}(k)$  は音源のスペクトル  $Z_j^{(i)}(k)$  と音源の到來方向  $\theta_i$  の HRTF の積であるため, 次式のようになる.

$$Z_{L,j}^{(i)}(k, m) = H_L(k, \theta_i)Z_j^{(i)}(k) \quad (3.11)$$

$$Z_{R,j}^{(i)}(k, m) = H_R(k, \theta_i)Z_j^{(i)}(k) \quad (3.12)$$

上記の式より, 式 (3.9) と式 (3.10) における時変係数  $Y_j$  を左右で共通のものとして扱うことができる. したがって, 次式が得られる.

$$\left[ X_L^{(i)}(k, m) \ X_R^{(i)}(k, m) \right] = \sum_j^{\rho} Y_j^{(i)}(m) \left[ Z_{L,j}^{(i)}(k) \ Z_{R,j}^{(i)}(k) \right] \quad (3.13)$$

以上の式変形により,  $i$  個目の音源の  $j$  番目のスペクトルは左右でペア化された状態で推定される. ISA では推定した  $Z_j(k)$  を部分空間ごとに分類する必要があり, 詳細は次の節で述べる.

最後に, 推定したスペクトログラムに対し逆短時間フーリエ変換を行い,  $i$  個目の音源  $s_i(n)$  を得る.

$$s_i(n) = STFT^{-1}\{X^{(i)}(k, m)\} \quad (3.14)$$

以上の手順は観測音源のスペクトログラム  $X(k, m)$  が定常であるという仮定のもとで行われている. 非定常なスペクトログラムの場合は,  $X(k, m)$  を, 定常であると仮定した短い時間  $\Delta M$  の間隔で切り出し上記の処理を行う. 分離されたスペクトログラムを滑らかにするため, 切り出し範囲は  $\Delta/2$  ずつオーバーラップする.

### 3.2 方向推定によるスペクトルの分類

従来手法における方向推定によるスペクトルの分類の処理について述べる.

ISA の後処理として, 推定した左右のスペクトル  $Z_{L,j}(k), Z_{R,j}(k)$  を部分空間ごとに分類しなければならない. 前節の式変形により, 同じ音源のスペクトルが左右でペアとなった状態で推定されることがわかっているので, スペクトルに畠み込まれた HRTF にもとづいて到来方向ごとにスペクトルを分類すれば, 同時に部分空間ごとにスペクトルを分類することができる.

$Z_{L,j}(k), Z_{R,j}(k)$  は音源のスペクトル  $Z_j(k)$  に到来方向  $\theta_i$  の HRTF が畠み込まれたものであるため,  $Z_j$  に畠み込まれている HRTF から得られる IPD や ILD を用いて到来方向を推定することができる. IPD は 1.5kHz 以上の周波数では位相が  $-\pi$  から  $\pi$  の範囲を超えることによる多義性が生じるが, 複数の周波数における IPD を用いることで回避できる.

従来手法では以下のように推定したスペクトルを部分空間ごとに分類する.

ISA により推定した  $Z_{R,j}(k)$  を  $Z_{L,j}(k)$  で除算することにより両耳間差を得る.

$$\frac{Z_{R,j}(k)}{Z_{L,j}(k)} = \frac{H_R(k, \theta_i)Z_j(k)}{H_L(k, \theta_i)Z_j(k)} \quad (3.15)$$

$$= \frac{H_R(k, \theta_i)}{H_L(k, \theta_i)} \quad (3.16)$$

スペクトル  $Z_j$  の IPD  $\psi_{obs}(k)$  は次のように計算できる.

$$\psi_{obs}(k) = \arg \left\{ \frac{Z_{R,j}(k)}{Z_{L,j}(k)} \right\} \quad (3.17)$$

HRTF データベースの IPD  $\psi(k, \theta)$  は次式で定義されている.

$$\psi(k, \theta) = \arg \left\{ \frac{H_R(k, \theta)}{H_L(k, \theta)} \right\} \quad (3.18)$$

次の評価関数  $EVL(\theta)$  が最小となる方向  $\theta$  が  $Z_j^{(i)}(k)$  の到来方向  $\theta_i$  であると推定する.

$$EVL(\theta) = \sum_{k=0}^{K-1} E(k) |\psi_{obs}(k) - \psi(k, \theta)| \quad (3.19)$$

$E(k)$  は雑音を除去するための重みであり, 推定された左右のスペクトログラムのパワーの平均である。式で表すと以下のようになる.

$$E(k) = \frac{|Z_{L,j}(k)| + |Z_{R,j}(k)|}{2} \quad (3.20)$$

到来方向の推定により,  $Z_j(k)$  を部分空間ごとに分類する. これにより式 (3.13) を得る.

最後に, 推定した音源のスペクトログラム  $X_L^{(i)}(k, m)$ ,  $X_R^{(i)}(k, m)$  に対し逆短時間フーリエ変換を行い,  $i$  個目の左右の分離音源  $\tilde{s}_{L,i}(n)$ ,  $\tilde{s}_{R,i}(n)$  をを得る.

$$\tilde{s}_{L,i}(n) = STFT^{-1}\{X_L^{(i)}(k, m)\} \quad (3.21)$$

$$\tilde{s}_{R,i}(n) = STFT^{-1}\{X_R^{(i)}(k, m)\} \quad (3.22)$$

### 3.3 まとめ

本章では提案手法の基礎となる独立部分空間分析やそれを用いた音源分離の際のスペクトルの分類の処理について説明した. ICA を拡張した手法である ISA は, 音源のスペクトログラムが複数のスペクトログラムの荷重和で表現されると考えることにより, 一つの観測音源から複数の音源を分離することができる手法であることを説明し, 左右の観測音源に同時に ISA を行うことで推定される成分の順序が不定である問題を解決した手法を説明した. また, 前述の手法では推定したスペクトログラムの IPD とデータベースの IPD との差の荷重和を用いて到来方向の推定及び部分空間への分類を行っていることを説明した.

## 第 4 章 提案手法

本章では提案手法について述べる。HRTF のデータベースとの比較に用いる角度ごとの特徴量として複素誤差の絶対値を用い、荷重にクロススペクトル用いることを提案する。まず、ISA により推定したスペクトルから得られる角度ごとの特徴量として複素誤差の絶対値が使用できることを示し、従来の IPD よりも特徴として優れている点について述べる。そして、比較の際の荷重としてクロススペクトルを用いる方法を示し、従来手法の荷重よりも優れている点について述べる。

### 4.1 複素誤差の絶対値

ISA で推定したスペクトログラムは、後処理として到来角度ごとに分類する必要がある。スペクトログラムには HRTF が含まれているため、HRTF を角度ごとの特徴量として利用できる。しかし推定したスペクトログラムには耳に到達する前の元の音源の情報が含まれているため、推定した左右のスペクトログラム同士で除算をすることで元の音源の情報を取り除く必要がある。以下に推定した左右のスペクトログラム  $X_R(k)$  と  $X_L(k)$  で除算することにより、元の音源のスペクトログラム  $Z(k)$  を除く式を示す。

$$\frac{Z_R(k)}{Z_L(k)} = \frac{H_R(k, \theta)Z(k)}{H_L(k, \theta)Z(k)} = \frac{H_R(k, \theta)}{H_L(k, \theta)} \quad (4.1)$$

( $k$ :周波数インデックス 1, 2, ..., K-1)

これにより到来方向  $\theta$  に依存する左右の HRTF  $H_R(k, \theta)$  と  $H_L(k, \theta)$  の商を得ることができる。データベースとの比較において、従来の手法では上記の商から偏角を計算し得られる IPD  $\psi_{obs}$  を用いており、以下にその評価関数  $EVL_a(\theta)$  を示す。

$$EVL_a(\theta) = \sum_{k=0}^{K-1} E(k) |\psi_{obs}(k) - \psi_{DB}(k, \theta)| \quad (4.2)$$

$\psi_{DB}(k, \theta)$  はデータベースにおける IPD である。また、 $E(k)$  は荷重であり、次の節で詳しく述べる。 $EVL_a(\theta)$  が最小となる方向  $\theta$  がそのスペクトログラムの到来方向として分類される。

しかし IPD は元の HRTF における位相情報しか含んでおらず、振幅に関する情報が切り捨てられている。一般的に、情報が似ているかどうか判断するための尺度である類似度は、考慮する情報が多いほど精度が高くなると考えられる。そのため、位相と同時に振幅情報も

比較に用いることで、従来の IPD を用いるより正確な到来方向への分類の精度が向上すると考えられる。

そこで本論文では、従来の IPD に代わり複素誤差の絶対値を提案する。以下に複素誤差の絶対値を用いた評価関数  $EVL_b(\theta)$  を示す。

$$EVL_b(\theta) = \sum_{k=0}^{K-1} E(k) \left| \frac{Z_R(k)}{Z_L(k)} - \frac{H_R(k, \theta)}{H_L(k, \theta)} \right| \quad (4.3)$$

従来手法同様、 $EVL_b(\theta)$  が最小となる方向  $\theta$  をそのスペクトログラムの到来方向として分類する。上記の式の絶対値内において、左の項は式 (4.1) に示すように、そのスペクトログラムの HRTF のみとなる。よってデータベースの HRTF との比較を行うことができる。HRTF は実部と虚部を持つ複素数で表されるため、絶対値内における差は複素数の差、すなわち複素誤差となる。IPD を計算する前の複素数の状態で差を取ることにより、振幅の情報もデータベースとの比較に用いることができる。よって提案手法である複素誤差の絶対値を用いることで、従来手法の IPD より到来方向への分類精度が高くなると考えられる。

## 4.2 クロススペクトル

到来方向を推定するための評価関数には、有意な周波数スペクトルのみを考慮するための荷重  $E(k)$  が用いられている。荷重により、有意なスペクトルほど評価関数に与える影響が大きくなる。従来の手法では、荷重  $E(k)$  は推定した左右のスペクトログラムのパワーの平均を用いている。以下にその式を示す。

$$E(k) = \frac{|Z_L(k)| + |Z_R(k)|}{2} \quad (4.4)$$

雑音が一定であるならば、パワーが大きい周波数スペクトルほど元の音源や HRTF のパワーが大きくなるため、その周波数スペクトルにおける雑音の比率が少ないという考え方に基づいて従来手法の荷重が決定されている。しかし従来手法であるパワーの平均では、例えば  $Z_L(k)$  のパワーが大きく  $Z_R(k)$  のパワーが小さいというような、左右でパワーの大きさが異なっている場合において、左右のパワーが一つに平均化されてしまうために、有意でない周波数スペクトルが評価関数に大きく影響したり、逆に有意なスペクトルであるのに評価関数への影響が小さくなってしまったりする可能性がある。

そこで本論文では、従来手法の左右のスペクトログラムのパワーの平均に代わり、クロススペクトルを提案する。以下にクロススペクトルによる荷重  $E(k)$  の式を示す。

$$E(k) = |Z_L(k)Z_R(k)^*| \quad (4.5)$$

ここで  $Z_R(k)^*$  は  $Z_R(k)$  の複素共役である。クロススペクトルを用いることで、左右でパワーの大きさが異なる場合は評価関数に与える影響が小さくなり、左右どちらも大きい場合のみ評価関数への影響が大きくなる。これにより、提案手法であるクロススペクトルを用いることで、従来手法のパワー平均を用いた場合に有意ではないにも関わらず評価関数への影響が大きかった周波数スペクトルを減らし、有意なスペクトルのみに荷重がかかり到来方向の分類精度が向上すると考えられる。

### 4.3 まとめ

本章では、角度ごとの特徴量と重みに複素誤差の絶対値とクロススペクトルを提案した。従来手法では角度ごとの特徴量として IPD を用いていたが、IPD は HRTF の位相情報であり、HRTF の振幅の情報を考慮していない。データベースとの比較には位相に加えて振幅も考慮することで到来方向の分類精度が向上すると考えられる。そこで提案手法として IPD を計算する前の HRTF の複素数をそのまま比較に用いる、複素誤差の絶対値を提案した。また、データベースとの比較に用いる荷重は、有意なスペクトルが評価関数への影響を大きくするためのものであり、従来は推定した左右のスペクトログラムのパワーの平均を用いていたが、左右のパワーが異なる場合において正しく重みを決定できない可能性があった。そこで提案手法としてクロススペクトルを提案し、左右のスペクトログラムのパワーがどちらも大きい場合のみを有意なスペクトルとして重み付けするようにした。これにより、従来手法より分類精度が向上すると考えられる。

# 第 5 章 シミュレーション

本章では角度ごとの特徴量に複素誤差の絶対値を、荷重にクロススペクトルを用いることの有効性について、従来手法における角度ごとの特徴量 IPD と荷重の左右のスペクトログラムのパワー平均、またそれぞれ特徴量と荷重の組み合わせを変えたシミュレーションを実行し、分離性能を比較することにより確認する。

## 5.1 実験条件と評価指標

本節ではシミュレーションの各条件と評価の指標について述べる。

### 5.1.1 実験条件

本シミュレーションでは、推定したスペクトログラムを到来方向に分類するための角度ごとの特徴量として IPD もしくは複素誤差の絶対値を使用し、特徴量にかかる荷重に推定した左右のスペクトログラムのパワー平均もしくはクロススペクトルを使用した。特徴量と荷重の組合せを変えてそれぞれシミュレーションを実行した。

また、シミュレーションには女性のスピーチ、男性のスピーチ、音楽の3つの音源から2つの音源を選択し、組み合わせを変えてシミュレーションを行った。1つ目の音源は頭部の正面、すなわち  $\theta=0^\circ$  を到来方向とし、2つ目の音源は  $\theta=30, 60, 90, \dots, 330^\circ$  の各角度の中で一つの角度を到来方向として、観測音源を生成した。観測音源は音源の到来方向により変化するため、2つ目の音源は  $30^\circ$  ごとの角度すべてでシミュレーションを行った。正面到来方向とする音源を  $s_1$ 、もう一つの音源を  $s_2$  としたときのシミュレーション実験図を図5.1に示す。また、元の音源から生成される観測音源  $x_L, x_R$  の式を、頭部インパルス応答  $h(n, \theta_i)$  を用いて以下に示す。

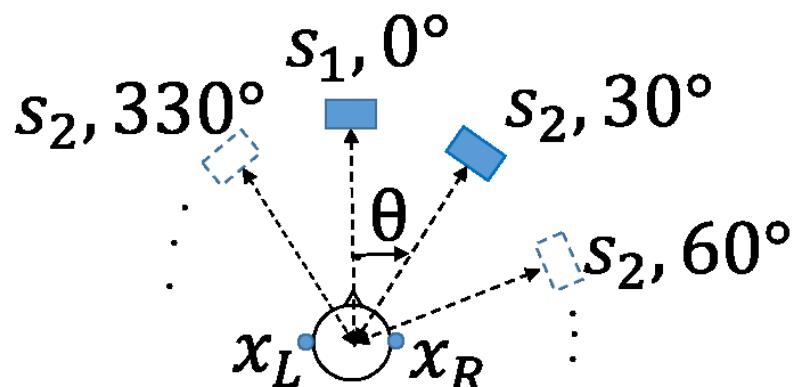
$$x_L(n) = h_L(n, 0^\circ) * s_1(n) + h_L(n, \theta) * s_2(n) \quad (5.1)$$

$$x_R(n) = h_R(n, 0^\circ) * s_1(n) + h_R(n, \theta) * s_2(n) \quad (5.2)$$

その他本シミュレーションで用いたパラメータを表5.1に示す。

表 5.1 シミュレーション条件

角度分解能	5°
各音源の時間(長さ)	5.0 s
サンプリング周波数	44.1 kHz
FT の窓関数の点数	1024
STFT のオーバーラップ点数	768
ISA の時間幅	87.1 ms



$[s_2$  の角度  $\theta : 30, 60, 90, 120, 150,$   
 $180, 210, 240, 270, 300, 330^\circ]$

図 5.1 シミュレーション実験図

### 5.1.2 評価指標

本シミュレーションでは分離性能の評価に相関係数を用いる。到來方向  $\theta_i$  から耳に到達した  $i$  個目の音源  $s_i$  を原音源と呼び、原音源と分離音源との相関係数を評価指数とする。原音源には角度  $\theta_i$  に応じた頭部インパルス応答  $h(n, \theta_i)$  が畳み込まれており、左右の原音源  $s_{L,i}(n), s_{R,i}(n)$  を式で表すと以下のようになる。

$$s_{L,i}(n) = h_L(n, \theta_i) * s_i(n) \quad (5.3)$$

$$s_{R,i}(n) = h_R(n, \theta_i) * s_i(n) \quad (5.4)$$

また、左右の分離音源を  $\tilde{s}_{L,i}(n), \tilde{s}_{R,i}(n)$  とすると、 $i$  個目の音源の相関係数  $C_i$  は以下の式で求められる。

$$C_{L,i} = \frac{\sum(s_{L,i} - E\{s_{L,i}\}) \sum(\tilde{s}_{L,i} - E\{\tilde{s}_{L,i}\})}{\sqrt{\sum(s_{L,i} - E\{s_{L,i}\})^2} \sqrt{\sum(\tilde{s}_{L,i} - E\{\tilde{s}_{L,i}\})^2}} \quad (5.5)$$

$$C_{R,i} = \frac{\sum(s_{R,i} - E\{s_{R,i}\}) \sum(\tilde{s}_{R,i} - E\{\tilde{s}_{R,i}\})}{\sqrt{\sum(s_{R,i} - E\{s_{R,i}\})^2} \sqrt{\sum(\tilde{s}_{R,i} - E\{\tilde{s}_{R,i}\})^2}} \quad (5.6)$$

ここで、 $E\{\cdot\}$  は期待値を表す。

相関係数は 0 から 1 の値で定義され、1 に近いほど原音源と分離音源の波形が似た音源となる。よって分離音源と原音源の相関係数が大きいほど、その分離音源を生成した手法の分離性能が高いとみなせる。

## 5.2 実験結果

本節では角度ごとの特徴量に複素誤差の絶対値を、荷重にクロススペクトルを用いることの有効性を測るために、従来における角度ごとの特徴量の IPD と荷重のパワー平均と合わせてそれぞれ特徴量と荷重の組み合わせを変えてシミュレーションを実行し、分離精度の比較を行う。ここで、角度ごとの特徴量に複素誤差の絶対値を、荷重にクロススペクトルを用いることを提案手法と称し、IPD とパワー平均を用いるものを従来手法と称する。また、シミュレーションに用いる音源の組み合わせを case1, case2 として表 5.2 に示す。なお、本シミュレーションでは正中面の判定を行った場合と行わなかった場合の組み合わせでもシミュレーションを行う。正中面の判定を行わない場合、到來方向の分類の際に  $0^\circ$  と  $180^\circ$  の区別がつかないため、 $180^\circ$  の角度に分類されるべきスペクトログラムがすべて  $0^\circ$  に分類される。よって正中面の判定を行わない場合、 $180^\circ$  の結果は存在しない。

表 5.2 音源の組み合わせ

	$s_1$	$s_2$
case1	女性スピーチ	音楽
case2	男性スピーチ	女性スピーチ

また、重みにクロススペクトルを用い、角度ごとの特徴量に従来手法の IPD を用いる組み合わせや、重みに従来のパワー平均を用いて角度ごとの特徴量に複素誤差の絶対値を用いる組み合わせでのシミュレーションは実行したが、一部の角度を除き従来手法の特徴量と重みの組み合わせ以下の相関係数であったため、結果の視認性を考慮して本章では示さない。これらの詳しい数値については付録にて本章の結果とともに示す。

### 5.2.1 case1

$\theta$  を変化させたときの相関係数を算出した結果を図 5.2 に示す。同図には観測音源と原音源の相関係数、従来手法により分離した音源と原音源の相関係数、提案手法により分離した音源と原音源の相関係数がそれぞれ示されている。図に示された相関係数は、左右の音源から得られた相関係数を平均したものである。図 5.2 より、提案手法と従来手法の相関係数が観測音源、つまり未処理の音源の相関係数を上回っている。これにより、観測音源から原音源へと分離ができていることがわかる。また、図 5.2(a) では  $90^\circ$ ,  $150^\circ$ ,  $210^\circ$  の角度を除き提案手法が従来手法の相関係数を上回っており、従来手法に比べて提案手法の相関係数が平均 0.020 向上している。図 5.2(b) では  $90^\circ$  の角度を除き提案手法が従来手法の相関係数を上回っており、提案手法が従来手法より平均 0.035 上回った。これらの結果から、図 5.2 全体では 20 点の角度のうち 15 点の角度で従来手法より改善されていることがわかる。

正中面判定を実行した場合の相関係数のグラフを図 5.3 に示す。図 5.3(a) より、正中面判定を行うことで図 5.2(a) よりも女性スピーチの相関係数が減少することが確認できる。これは正中面判定を行うことにより、 $0^\circ$  の角度に分類されていた多数のスペクトログラムが、 $180^\circ$  の角度に分類されなおされたためと考えられる。図 5.2(a) と同様に、 $90^\circ$ ,  $150^\circ$ ,  $210^\circ$  の角度以外では従来手法の相関係数を上回っており、平均 0.015 向上している。しかし、 $120^\circ$ ,  $240^\circ$  の角度では、正中面判定を行うことによって、観測音源の相関係数より低い値となってしまっており、これらの角度ではうまく原音源へ分離できていないことがわかる。また図 5.3(a), (b) のどちらの図でも  $180^\circ$  の角度における相関係数は提案手法と従来手法で全く同

一の値を取り、観測音源の相関係数を下回っている。観測音源の相関係数を下回っているため、観測音源から原音源への分離ができていないことがわかるが、提案手法と従来手法の相関係数が同一の値であることから、この角度における分離を達成するには正中面判定の精度を上げる必要があることがわかる。

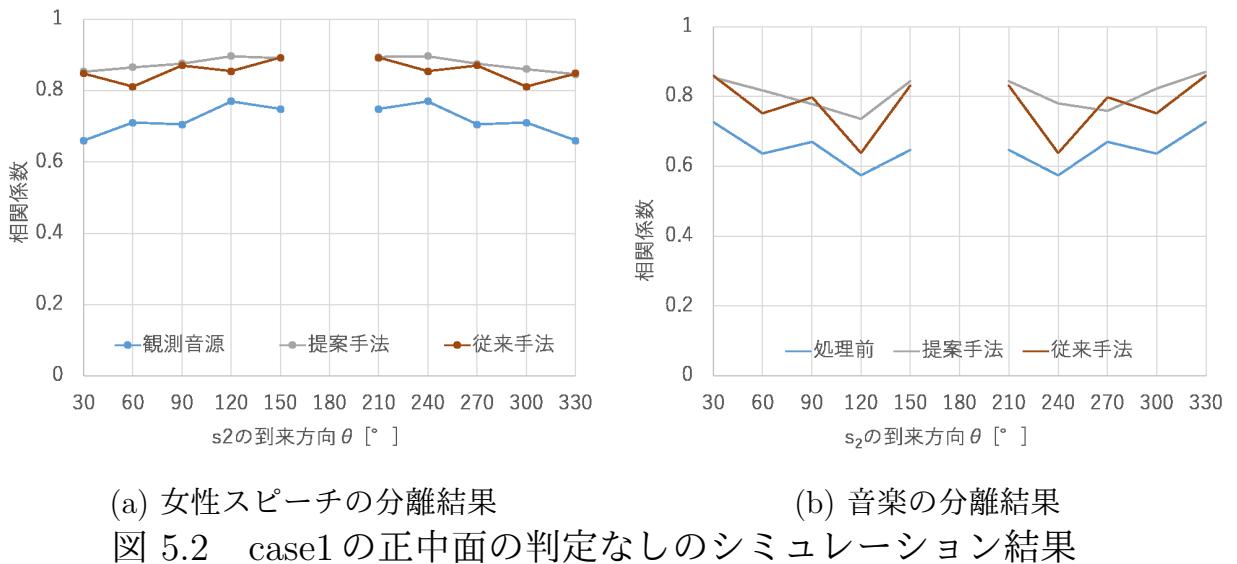
それぞれの相関係数の数値データを付録 A.1 に示す。

### 5.2.2 case2

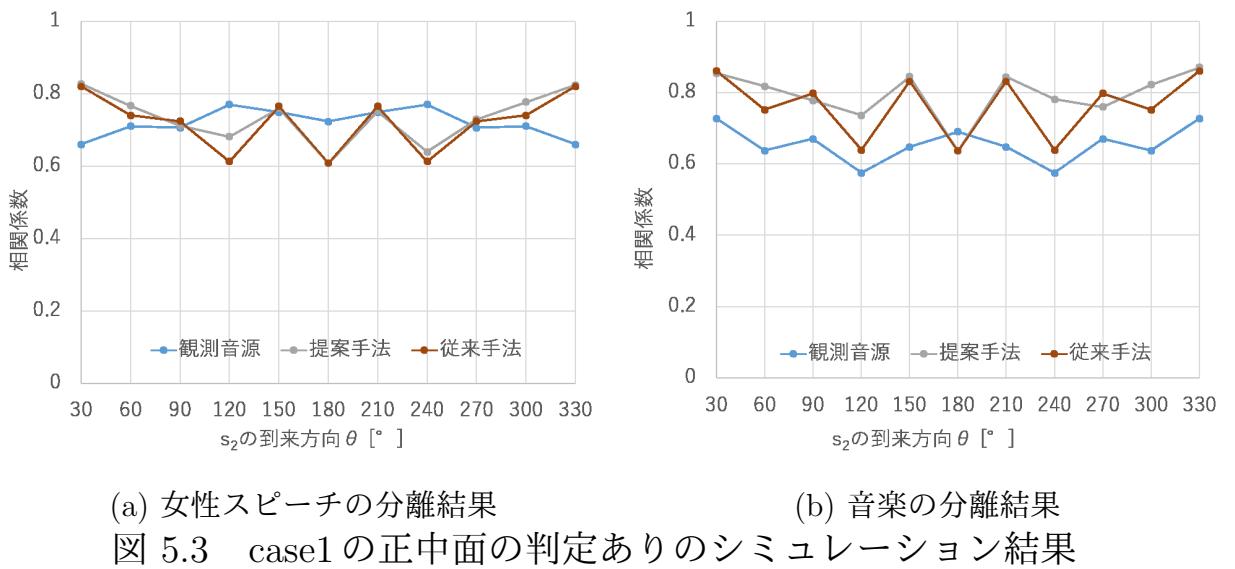
$\theta$  を変化させたときの相関係数を算出した結果を図 5.4 に示す。同図より、case1 同様提案手法と従来手法の相関係数が観測音源の相関係数を上回っており、正常に分離できていることがわかる。また、図 5.4(a) では  $270^\circ$  を除いたすべての角度で提案手法の相関係数が従来手法の相関係数を上回っており、従来手法に比べ提案手法の相関係数が平均 0.016 程度向上している。図 5.4(b) では  $60^\circ$ ,  $120^\circ$ ,  $270^\circ$  の角度を除いて提案手法の相関係数が従来手法のものより上回った。図 5.4(b) の相関係数の差を平均すると、提案手法の相関係数のほうが従来手法より 0.012 程度上回っていることがわかる。これらの結果から、図 5.4 全体では 20 点の角度のうち 16 点の角度で従来手法より改善されていることがわかった。

正中面判定を実行した場合の相関係数のグラフを図 5.5 に示す。図 5.5(a) より、case1 同様、正中面判定を行うことで図 5.4(a) よりも男性スピーチの相関係数が減少していることが確認できる。図 5.5(a) では  $90^\circ$  と  $150^\circ$  の角度を除いて従来手法の相関係数を上回っており、平均 0.010 向上している。しかし、 $30^\circ$  と  $330^\circ$  以外の角度では、正中面判定を行うことによって、観測音源の相関係数より低い値となっており、ほぼすべての角度でうまく原音源へ分離できていないことがわかる。また図 5.5(a), (b) のどちらの図でも case1 同様  $180^\circ$  の角度における相関係数は提案手法と従来手法で全く同一の値を取り、観測音源の相関係数を下回っている。

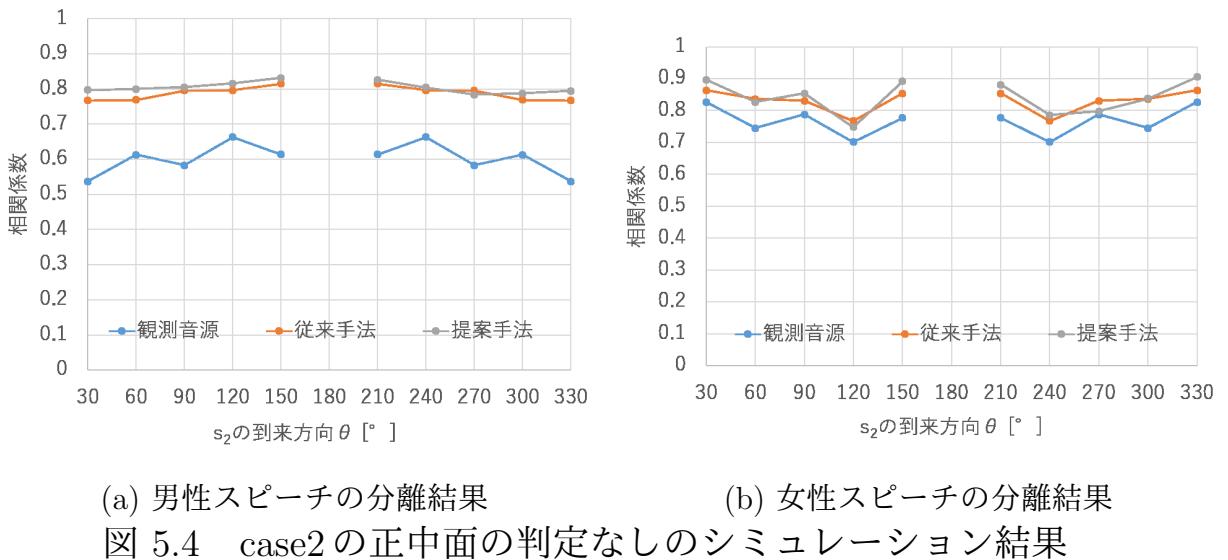
それぞれの相関係数の数値データを付録 A.2 に示す。



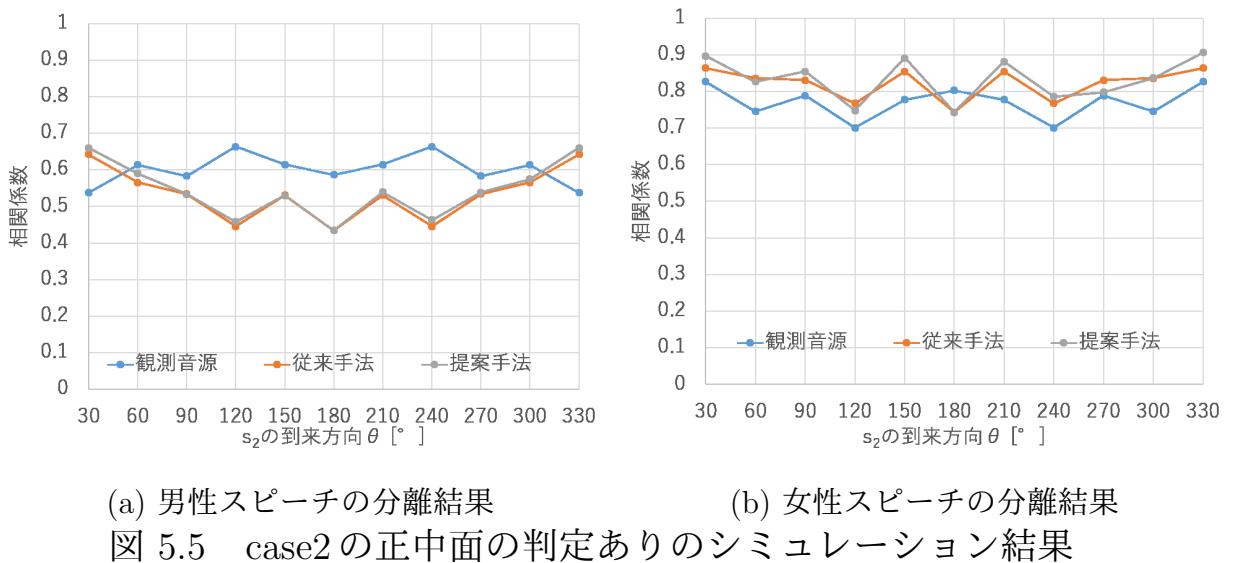
(a) 女性スピーチの分離結果  
(b) 音楽の分離結果  
図 5.2 case1 の正中面の判定なしのシミュレーション結果



(a) 女性スピーチの分離結果  
(b) 音楽の分離結果  
図 5.3 case1 の正中面の判定ありのシミュレーション結果



(a) 男性スピーチの分離結果  
(b) 女性スピーチの分離結果  
図 5.4 case2 の正中面の判定なしのシミュレーション結果



(a) 男性スピーチの分離結果  
(b) 女性スピーチの分離結果  
図 5.5 case2 の正中面の判定ありのシミュレーション結果

### 5.3 考察

case1 と case2 のシミュレーションから、全体的に従来手法よりも提案手法のほうが分離精度が向上していることがわかった。しかし、正中面判定を実行したシミュレーションの結果では、観測音源、つまり未処理の音源よりも相関係数が下がってしまっており、うまく分離ができていない角度が存在することがわかる。 $0^\circ$  の角度から到來した音源 (case1 であれば女性のスピーチ, case2 であれば男性のスピーチ) の相関係数が正中面判定を実行しない場合より低くなるのは、 $0^\circ$  の角度に分類されていたスペクトログラムが、 $180^\circ$  の角度に分類されなおされたためと考えられる。 $\theta$  が  $180^\circ$  以外の角度である場合、 $180^\circ$  の角度にスペクトログラムが分類されることはないので、正中面判定の精度を上げることで、正中面判定を行っていないときの相関係数と同じ値まで相関係数を向上できる。また、角度ごとに分類されたスペクトログラムの数がどうなっているか調べたものを表 5.3, 表 5.4 に示す。表は原音源  $s_2$  が角度  $\theta$  から到來した際に、正面の音源  $s_1$  と推定されたスペクトログラムの数、 $\theta$  から到來した音源  $s_2$  と推定されたスペクトログラムの数、それ以外の方向だと推定されたスペクトログラムの数をそれぞれ示している。スペクトログラムの合計は 1710 となる。表によると、従来手法のほうが提案手法よりも目的とする方向以外に分類したスペクトログラムの数が少なくなっている。一方で前述したように分離精度は提案手法が従来手法を上回っている。これには 2 つの推測が立てられる。まず 1 つめは、目的とする音源の角度以外に分類されたスペクトログラムは無音などの無意味な情報である可能性である。無音のスペクトログラムならばどの角度に分類されても、目的音源の相関係数が下がることはないと想定される。しかし正確にスペクトログラムの分類が可能ならば、目的とする音源の方向以外にスペクトログラムが分類されるのは本来ならありえないことである。また、無音のスペクトログラムは相関係数が下がらない要因にはなっても上がる要因にはならない。そこでもう一つの可能性が考えられる。それは従来手法によって間違って分類されたスペクトログラムが、提案手法によって正しく分類されたという可能性である。これにより分離精度が向上したと思われる。よって角度ごとの特徴量に複素誤差の絶対値を用いることと、荷重にクロススペクトルを用いることは有効である。しかし目的とする音源の方向以外に分類されてしまうスペクトログラムが多いため、それらのスペクトログラムを目的とする音源がある方向に分類できるように改善することでさらなる分離精度の向上が見込める。

表 5.3 case1 の正中面判定なしにおけるスペクトログラムの分類数

$\theta$	女性 ( $0^\circ$ )		音楽 ( $\theta$ )		その他	
	提案	従来	提案	従来	提案	従来
30	480	614	884	1047	346	49
60	713	711	730	969	267	30
90	668	706	795	877	247	127
120	777	808	621	789	312	113
150	548	699	932	944	230	67
210	509	699	968	944	233	67
240	551	808	867	789	292	113
270	455	706	862	877	393	127
300	444	711	1051	969	215	30
330	387	614	1155	1047	168	49

表 5.4 case2 の正中面判定なしにおけるスペクトログラムの分類数

$\theta$	男性 ( $0^\circ$ )		女性 ( $\theta$ )		その他	
	提案	従来	提案	従来	提案	従来
30	724	830	697	767	289	113
60	894	899	591	767	225	44
90	828	959	641	622	241	129
120	968	995	430	612	312	103
150	773	899	719	735	218	76
210	737	899	745	735	228	76
240	795	995	564	612	351	103
270	696	959	532	622	482	129
300	721	899	742	767	247	44
330	672	830	866	767	172	113

## 5.4 まとめ

本章では、提案手法と従来手法の角度ごとの特徴量と重みの組み合わせを変えてシミュレーションを実行し、分離性能の比較を行った。その結果、角度ごとの特徴量と重みの組み合わせは提案手法である複素誤差の絶対値とクロススペクトルの組み合わせが最も分離性能が高いことがわかった。一方で正中面判定を行うと、未処理の音源との相関係数を上回れずうまく分離できていなかったり、角度ごとに分類されたスペクトログラムの数を調べると提案手法より従来手法のほうが目的とする方向に多くのスペクトログラムを分類できていることがわかった。提案手法のほうが相関係数が高かったため、従来手法では間違った角度に分類されていたスペクトログラムを正しく分類できるようになったと考えられるが、目的とする音源がある方向以外に分類されてしまったスペクトログラムを、目的とする音源がある方向に分類できるよう改善することでさらなる分離精度の向上が見込める。

## 第 6 章 結論

本論文では、ISA を両耳聴モデルに適用した分離手法において、到來方向ごとにスペクトログラムを分類する処理の改善のため、角度ごとの特徴量に複素誤差の絶対値を用いることと、データベースとの比較に用いる荷重にクロススペクトルを用いることについて提案した。従来手法では角度ごとの特徴量に IPD を用いていたが、IPD は振幅の情報を考慮しない。そのため、複素誤差の絶対値で位相と振幅の両方を考慮することで、到來方向の分類精度が向上すると考えた。また、データベースとの比較に用いる荷重は従来では推定した左右のスペクトログラムのパワー平均のため、左右のスペクトログラムでパワーが異なる場合に、正しく重みを決定できない可能性があった。そこでクロススペクトルを用いることで、左右のスペクトログラムのパワーのどちらも大きい場合のみ有意な周波数となるよう重みを決定することができる。これにより、分類精度が向上すると考えた。

提案した角度ごとの特徴量としての複素誤差の絶対値や、荷重としてのクロススペクトルが有効であることを確かめるため、分離精度の指標として相関係数を用いてシミュレーションを行った。シミュレーションは元となる音源の組み合わせや到來方向、正中面を判定するか否かなどを変えてそれぞれ実行した。

女性のスピーチと音楽の組み合わせで行った case1 のシミュレーションでは、少数の角度の組み合わせを除き、提案手法のほうが従来手法よりも分離精度が高くなった。相関係数では、提案手法により従来手法より女性では 0.015～0.020 の上昇が、音楽では 0.035 の上昇が確認できた。また、男性のスピーチと女性のスピーチを組み合わせて行った case2 のシミュレーションにおいても同様に、少数の角度を除いて提案手法のほうが従来手法よりも分離精度が高くなかった。相関係数では男性のほうが 0.010～0.016 の上昇が、女性のほうは 0.012 程度の上昇が確認できた。角度ごとに分類されたスペクトログラムの数を比較すると、提案手法の目的とする音源の方向以外の角度に分類されたスペクトログラムの数が従来手法のそれよりも多いことが示された。スペクトログラムの数に対し、分離精度は向上しているため、従来手法では間違えて分類されていたスペクトログラムを正しい角度に分類できたと推測でき、提案手法の有効性を確認できた。

## 付録A シミュレーション結果

### A.1 シミュレーション case1 の結果

表 A.1, 表 A.2 に case1 の正中面判定なしの女性のスピーチのシミュレーション結果を示す。また、表 A.3, 表 A.4 に case1 の正中面判定なしの音楽のシミュレーション結果を示し、表 A.5, 表 A.6 に case1 の正中面判定ありの女性のスピーチのシミュレーション結果を、表 A.7, 表 A.8 に case1 の正中面判定ありの音楽のシミュレーション結果を示す。

表 A.1 case1 の正中面判定なしにおける女性スピーチの相関係数 1.

$\theta$	観測音源			提案手法			従来手法		
	左	右	平均	左	右	平均	左	右	平均
30	0.808	0.512	0.660	0.890	0.816	0.853	0.900	0.796	0.848
60	0.915	0.505	0.710	0.939	0.793	0.866	0.914	0.709	0.811
90	0.858	0.554	0.706	0.925	0.827	0.876	0.927	0.815	0.871
120	0.939	0.602	0.770	0.954	0.840	0.897	0.942	0.767	0.854
150	0.852	0.645	0.749	0.913	0.870	0.891	0.923	0.862	0.892
210	0.645	0.852	0.749	0.873	0.916	0.895	0.862	0.923	0.892
240	0.602	0.939	0.770	0.856	0.938	0.897	0.767	0.942	0.854
270	0.554	0.858	0.706	0.839	0.911	0.875	0.815	0.927	0.871
300	0.505	0.915	0.710	0.814	0.907	0.861	0.709	0.914	0.811
330	0.512	0.808	0.660	0.814	0.877	0.846	0.796	0.900	0.848

表 A.2 case1 の正中面判定なしにおける女性スピーチの相関係数 2.

$\theta$	IPD&クロススペクトル			複素誤差の絶対値&パワー平均		
	左	右	平均	左	右	平均
30	0.882	0.795	0.839	0.897	0.793	0.845
60	0.926	0.771	0.848	0.939	0.704	0.822
90	0.914	0.820	0.867	0.923	0.796	0.859
120	0.941	0.834	0.888	0.953	0.735	0.844
150	0.914	0.870	0.892	0.918	0.861	0.889
210	0.870	0.914	0.892	0.861	0.900	0.880
240	0.834	0.941	0.888	0.835	0.891	0.863
270	0.820	0.914	0.867	0.813	0.862	0.837
300	0.771	0.926	0.848	0.781	0.838	0.810
330	0.795	0.882	0.839	0.783	0.828	0.806

表 A.3 case1 の正中面判定なしにおける音楽の相関係数 1.

$\theta$	観測音源			提案手法			従来手法		
	左	右	平均	左	右	平均	左	右	平均
30	0.600	0.855	0.728	0.825	0.884	0.854	0.815	0.906	0.861
60	0.409	0.865	0.637	0.762	0.873	0.817	0.638	0.865	0.751
90	0.506	0.835	0.670	0.757	0.798	0.778	0.747	0.848	0.798
120	0.352	0.798	0.575	0.684	0.788	0.736	0.543	0.734	0.639
150	0.535	0.760	0.647	0.817	0.871	0.844	0.793	0.870	0.832
210	0.760	0.535	0.647	0.869	0.818	0.844	0.870	0.793	0.832
240	0.798	0.352	0.575	0.859	0.703	0.781	0.734	0.543	0.639
270	0.835	0.506	0.670	0.786	0.732	0.759	0.848	0.747	0.798
300	0.865	0.409	0.637	0.914	0.730	0.822	0.865	0.638	0.751
330	0.855	0.600	0.728	0.914	0.826	0.870	0.906	0.815	0.861

表 A.4 case1 の正中面判定なしにおける音楽の相関係数 2.

$\theta$	IPD&クロススペクトル			複素誤差の絶対値&パワー平均		
	左	右	平均	左	右	平均
30	0.799	0.861	0.830	0.701	0.788	0.744
60	0.632	0.791	0.712	0.612	0.698	0.655
90	0.682	0.732	0.707	0.695	0.794	0.745
120	0.598	0.710	0.654	0.552	0.644	0.598
150	0.735	0.794	0.765	0.790	0.871	0.830
210	0.794	0.735	0.765	0.863	0.764	0.813
240	0.710	0.598	0.654	0.885	0.607	0.746
270	0.732	0.682	0.707	0.872	0.712	0.792
300	0.791	0.632	0.712	0.929	0.622	0.776
330	0.861	0.799	0.830	0.926	0.792	0.859

表 A.5 case1 の正中面判定ありにおける女性スピーチの相関係数 1.

$\theta$	観測音源			提案手法			従来手法		
	左	右	平均	左	右	平均	左	右	平均
30	0.808	0.512	0.660	0.863	0.791	0.827	0.870	0.771	0.820
60	0.915	0.505	0.710	0.828	0.706	0.767	0.821	0.659	0.740
90	0.858	0.554	0.706	0.750	0.672	0.711	0.767	0.682	0.724
120	0.939	0.602	0.770	0.726	0.636	0.681	0.685	0.541	0.613
150	0.852	0.645	0.749	0.780	0.740	0.760	0.794	0.736	0.765
180	0.723	0.723	0.723	0.608	0.608	0.608	0.608	0.608	0.608
210	0.645	0.852	0.749	0.732	0.771	0.751	0.736	0.794	0.765
240	0.602	0.939	0.770	0.612	0.666	0.639	0.541	0.685	0.613
270	0.554	0.858	0.706	0.702	0.756	0.729	0.682	0.767	0.724
300	0.505	0.915	0.710	0.741	0.814	0.777	0.659	0.821	0.740
330	0.512	0.808	0.660	0.793	0.855	0.824	0.771	0.870	0.820

表 A.6 case1 の正中面判定ありにおける女性スピーチの相関係数 2.

$\theta$	IPD&クロススペクトル			複素誤差の絶対値&パワー平均		
	左	右	平均	左	右	平均
30	0.860	0.776	0.818	0.868	0.769	0.818
60	0.823	0.692	0.757	0.829	0.635	0.732
90	0.757	0.682	0.720	0.736	0.636	0.686
120	0.680	0.598	0.639	0.687	0.522	0.605
150	0.781	0.740	0.760	0.784	0.731	0.757
180	0.608	0.608	0.608	0.608	0.608	0.608
210	0.740	0.781	0.760	0.719	0.755	0.737
240	0.598	0.680	0.639	0.603	0.643	0.623
270	0.682	0.757	0.720	0.669	0.707	0.688
300	0.692	0.823	0.757	0.706	0.753	0.729
330	0.776	0.860	0.818	0.760	0.804	0.782

表 A.7 case1 の正中面判定ありにおける音楽の相関係数 1.

$\theta$	観測音源			提案手法			従来手法		
	左	右	平均	左	右	平均	左	右	平均
30	0.600	0.855	0.728	0.825	0.884	0.854	0.815	0.906	0.861
60	0.409	0.865	0.637	0.762	0.873	0.817	0.638	0.865	0.751
90	0.506	0.835	0.670	0.757	0.798	0.778	0.747	0.848	0.798
120	0.352	0.798	0.575	0.684	0.788	0.736	0.543	0.734	0.639
150	0.535	0.760	0.647	0.817	0.871	0.844	0.793	0.870	0.832
180	0.690	0.690	0.690	0.636	0.636	0.636	0.636	0.636	0.636
210	0.760	0.535	0.647	0.869	0.818	0.844	0.870	0.793	0.832
240	0.798	0.352	0.575	0.859	0.703	0.781	0.734	0.543	0.639
270	0.835	0.506	0.670	0.786	0.732	0.759	0.848	0.747	0.798
300	0.865	0.409	0.637	0.914	0.730	0.822	0.865	0.638	0.751
330	0.855	0.600	0.728	0.914	0.826	0.870	0.906	0.815	0.861

表 A.8 case1 の正中面判定ありにおける音楽の相関係数 2.

$\theta$	IPD&クロススペクトル			複素誤差の絶対値&パワー平均		
	左	右	平均	左	右	平均
30	0.799	0.861	0.830	0.701	0.788	0.744
60	0.632	0.791	0.712	0.612	0.698	0.655
90	0.682	0.732	0.707	0.695	0.794	0.745
120	0.598	0.710	0.654	0.552	0.644	0.598
150	0.735	0.794	0.765	0.790	0.871	0.830
180	0.636	0.636	0.636	0.636	0.636	0.636
210	0.794	0.735	0.765	0.863	0.764	0.813
240	0.710	0.598	0.654	0.885	0.607	0.746
270	0.732	0.682	0.707	0.872	0.712	0.792
300	0.791	0.632	0.712	0.929	0.622	0.776
330	0.861	0.799	0.830	0.926	0.792	0.859

## A.2 シミュレーション case2 の結果

表 A.9, 表 A.10 に case2 の正中面判定なしの男性のスピーチのシミュレーション結果を示す。また、表 A.11, 表 A.12 に case2 の正中面判定なしの女性のスピーチのシミュレーション結果を示し、表 A.13, 表 A.14 に case2 の正中面判定ありの男性のスピーチのシミュレーション結果を、表 A.15, 表 A.16 に case2 の正中面判定ありの女性のスピーチのシミュレーション結果を示す。

表 A.9 case2 の正中面判定なしにおける男性スピーチの相関係数 1.

$\theta$	観測音源			提案手法			従来手法		
	左	右	平均	左	右	平均	左	右	平均
30	0.677	0.398	0.537	0.828	0.768	0.798	0.811	0.724	0.767
60	0.828	0.398	0.613	0.872	0.717	0.800	0.841	0.697	0.769
90	0.734	0.432	0.583	0.845	0.766	0.806	0.860	0.732	0.796
120	0.861	0.467	0.664	0.888	0.744	0.816	0.868	0.726	0.797
150	0.724	0.504	0.614	0.849	0.814	0.832	0.845	0.785	0.815
210	0.504	0.724	0.614	0.809	0.843	0.826	0.785	0.845	0.815
240	0.467	0.861	0.664	0.751	0.859	0.805	0.726	0.868	0.797
270	0.432	0.734	0.583	0.749	0.819	0.784	0.732	0.860	0.796
300	0.398	0.828	0.613	0.738	0.837	0.788	0.697	0.841	0.769
330	0.398	0.677	0.537	0.770	0.819	0.794	0.724	0.811	0.767

表 A.10 case2 の正中面判定なしにおける男性スピーチの相関係数 2.

$\theta$	IPD&クロススペクトル			複素誤差の絶対値&パワー平均		
	左	右	平均	左	右	平均
30	0.830	0.759	0.794	0.797	0.714	0.756
60	0.855	0.724	0.790	0.861	0.610	0.735
90	0.844	0.734	0.789	0.827	0.675	0.751
120	0.873	0.743	0.808	0.871	0.653	0.762
150	0.843	0.807	0.825	0.814	0.758	0.786
210	0.807	0.843	0.825	0.777	0.817	0.797
240	0.743	0.873	0.808	0.753	0.803	0.778
270	0.734	0.844	0.789	0.737	0.777	0.757
300	0.724	0.855	0.790	0.734	0.770	0.752
330	0.759	0.830	0.794	0.733	0.763	0.748

表 A.11 case2 の正中面判定なしにおける女性スピーチの相関係数 1.

$\theta$	観測音源			提案手法			従来手法		
	左	右	平均	左	右	平均	左	右	平均
30	0.737	0.917	0.827	0.867	0.926	0.897	0.823	0.904	0.863
60	0.568	0.923	0.745	0.751	0.901	0.826	0.737	0.935	0.836
90	0.668	0.908	0.788	0.807	0.902	0.854	0.788	0.873	0.831
120	0.514	0.888	0.701	0.662	0.835	0.748	0.659	0.876	0.768
150	0.696	0.859	0.778	0.860	0.923	0.892	0.814	0.895	0.854
210	0.859	0.696	0.778	0.910	0.852	0.881	0.895	0.814	0.854
240	0.888	0.514	0.701	0.876	0.697	0.786	0.876	0.659	0.768
270	0.908	0.668	0.788	0.830	0.765	0.798	0.873	0.788	0.831
300	0.923	0.568	0.745	0.932	0.742	0.837	0.935	0.737	0.836
330	0.917	0.737	0.827	0.940	0.871	0.906	0.904	0.823	0.863

表 A.12 case2 の正中面判定なしにおける女性スピーチの相関係数 2.

$\theta$	IPD&クロススペクトル			複素誤差の絶対値&パワー平均		
	左	右	平均	左	右	平均
30	0.816	0.871	0.843	0.810	0.890	0.850
60	0.752	0.918	0.835	0.635	0.816	0.725
90	0.717	0.777	0.747	0.758	0.858	0.808
120	0.665	0.847	0.756	0.563	0.771	0.667
150	0.784	0.851	0.818	0.841	0.919	0.880
210	0.851	0.784	0.818	0.911	0.839	0.875
240	0.847	0.665	0.756	0.928	0.681	0.804
270	0.777	0.717	0.747	0.899	0.801	0.850
300	0.918	0.752	0.835	0.958	0.718	0.838
330	0.871	0.816	0.843	0.956	0.856	0.906

表 A.13 case2 の正中面判定ありにおける男性スピーチの相関係数 1.

$\theta$	観測音源			提案手法			従来手法		
	左	右	平均	左	右	平均	左	右	平均
30	0.677	0.398	0.537	0.681	0.638	0.660	0.674	0.611	0.642
60	0.828	0.398	0.613	0.639	0.541	0.590	0.610	0.520	0.565
90	0.734	0.432	0.583	0.558	0.509	0.534	0.575	0.493	0.534
120	0.861	0.467	0.664	0.485	0.430	0.458	0.474	0.416	0.445
150	0.724	0.504	0.614	0.540	0.520	0.530	0.551	0.510	0.530
180	0.586	0.586	0.586	0.435	0.435	0.435	0.435	0.435	0.435
210	0.504	0.724	0.614	0.527	0.551	0.539	0.510	0.551	0.530
240	0.467	0.861	0.664	0.448	0.477	0.463	0.416	0.474	0.445
270	0.432	0.734	0.583	0.524	0.552	0.538	0.493	0.575	0.534
300	0.398	0.828	0.613	0.550	0.599	0.575	0.520	0.610	0.565
330	0.398	0.677	0.537	0.642	0.677	0.660	0.611	0.674	0.642

表 A.14 case2 の正中面判定ありにおける男性スピーチの相関係数 2.

$\theta$	IPD&クロススペクトル			複素誤差の絶対値&パワー平均		
	左	右	平均	左	右	平均
30	0.686	0.630	0.658	0.658	0.599	0.628
60	0.604	0.538	0.571	0.638	0.435	0.536
90	0.558	0.509	0.534	0.550	0.433	0.492
120	0.481	0.452	0.466	0.477	0.352	0.415
150	0.559	0.537	0.548	0.548	0.507	0.528
180	0.435	0.435	0.435	0.435	0.435	0.435
210	0.537	0.559	0.548	0.510	0.543	0.527
240	0.452	0.481	0.466	0.413	0.446	0.430
270	0.509	0.558	0.534	0.451	0.476	0.464
300	0.538	0.604	0.571	0.532	0.565	0.549
330	0.630	0.686	0.658	0.622	0.648	0.635

表 A.15 case2 の正中面判定ありにおける女性スピーチの相関係数1.

$\theta$	観測音源			提案手法			従来手法		
	左	右	平均	左	右	平均	左	右	平均
30	0.737	0.917	0.827	0.867	0.926	0.897	0.823	0.904	0.863
60	0.568	0.923	0.745	0.751	0.901	0.826	0.737	0.935	0.836
90	0.668	0.908	0.788	0.807	0.902	0.854	0.788	0.873	0.831
120	0.514	0.888	0.701	0.662	0.835	0.748	0.659	0.876	0.768
150	0.696	0.859	0.778	0.860	0.923	0.892	0.814	0.895	0.854
180	0.803	0.803	0.803	0.743	0.743	0.743	0.743	0.743	0.743
210	0.859	0.696	0.778	0.910	0.852	0.881	0.895	0.814	0.854
240	0.888	0.514	0.701	0.876	0.697	0.786	0.876	0.659	0.768
270	0.908	0.668	0.788	0.830	0.765	0.798	0.873	0.788	0.831
300	0.923	0.568	0.745	0.932	0.742	0.837	0.935	0.737	0.836
330	0.917	0.737	0.827	0.940	0.871	0.906	0.904	0.823	0.863

表 A.16 case2 の正中面判定ありにおける女性スピーチの相関係数2.

$\theta$	IPD&クロススペクトル			複素誤差の絶対値&パワー平均		
	左	右	平均	左	右	平均
30	0.816	0.871	0.843	0.810	0.890	0.850
60	0.752	0.918	0.835	0.635	0.816	0.725
90	0.717	0.777	0.747	0.758	0.858	0.808
120	0.665	0.847	0.756	0.563	0.771	0.667
150	0.784	0.851	0.818	0.841	0.919	0.880
180	0.743	0.743	0.743	0.743	0.743	0.743
210	0.851	0.784	0.818	0.911	0.839	0.875
240	0.847	0.665	0.756	0.928	0.681	0.804
270	0.777	0.717	0.747	0.899	0.801	0.850
300	0.918	0.752	0.835	0.958	0.718	0.838
330	0.871	0.816	0.843	0.956	0.856	0.906

## 謝辞

本論文の作成にあたり、終始適切なご指導をして頂いた、杉田泰則准教授に深く感謝いたします。また、論文の審査において、的確なご指示を頂きました、本学電気系岩橋政宏教授ならびに圓道知博准教授に感謝いたします。さらに、日頃お世話になりました信号処理応用研究室の皆様に感謝いたします。

最後に、勉学に励む機会を与え様々な面において支えていただいた両親に深く感謝いたします。

平成30年2月

## 参考文献

- [1] 澤田 宏, 荒木 章子, 牧野 昭二, ”音源分離技術の最新動向”, 電子情報通信学会学会誌, 91(4), 292–296, 2008
- [2] H. Sawada, S. Araki and S. Makino, “Underdetermined convolutive blind source separation via frequency bin-wise clustering and permutation alignment,” IEEE Trans. Audio, Speech, and Language Processing, vol. 19, no. 3, pp. 516–527, Mar., 2011.
- [3] T. Otsuka, K. Ishiguro, H. Sawada and H. G. Okuno, “Bayesian nonparametrics for microphone array processing,” IEEE/ACM Trans.Audio, Speech, and Language Processing, vol. 22, no. 2, pp. 493–504, Feb., 2014.
- [4] H. Saruwatari, T. Kawamura, T. Nishikawa, A. Lee and K. Shikano, “Blind source separation based on a fast-convergence algorithm combining ICA and beamforming,” IEEE Trans. Audio, Speech, and Language Processing, vol. 14, no. 2, pp. 666–678, Mar., 2006.
- [5] H. Sawada, S. Araki, S. Makino, “MLSP 2007 data analysys competition: Frequency-domain blind source separation for convolutive mixtures of speech/audio signals,” IEEE International Workshop on Machine Learning for Signal Processing, pp. 45–50, 2007.
- [6] A. Hyvärinen, J. Karhunen, E. Oja, “詳解 独立成分分析,” 東京大学出版局, 2005.
- [7] A. Hyvärinen and P. Hoyer, “Emergence of phase and shift invariant features by decomposition of natural images into independent feature subspaces,” Neural Computation, vol. 12, no. 7, pp. 1705-1720, July, 2000.
- [8] M. A. Casey and A. Westner, “Separation of mixed audio sources by independent subspace analysis,” Proc. International Computer Music Conference, pp. 154–161, 2000.
- [9] K. Iida, M. Itoh, A. Itagaki and M. Morimoto, “Median plane localization using a parametric model of the head-related transfer function based on spectral cues,” Applied Acoustics, vol. 68, no. 8, pp. 835–850, Aug., 2007

- [10] 飯田一博, 森本政之, ”空間音響学”, コロナ社, 2010.
- [11] MIT Media Lab., “HRTF of a KEMAR Dummy-Head,”  
<http://sound.media.mit.edu/resources/KEMAR.html>.
- [12] K. Iida, M. Itoh, A. Itagaki and M. Morimoto, “Median plane localization using a parametric model of the head-related transfer function based on spectral cues,” Applied Acoustics, vol. 68, no. 8, pp. 835–850, Aug., 2007
- [13] M. A. Casey and A. Westner, “Separation of mixed audio sources by independent subspace analysis,” Proc. International Computer Music Conference, pp. 154–161, 2000.
- [14] 西川直希, ”頭部伝達関数に基づく水平面・正中面における音源の方向推定及び分離手法に関する研究”, 長岡技術科学大学大学院工学研究科修士論文, 2015